Who is Who in the Psalms?

Erwich, C.M.

2021

document version
Publisher's PDF, also known as Version of record

Summary

In the Psalms it is a challenge to identify 'who is who', 'who speaks where' and 'who does what to whom' in the text. The study presented in this dissertation addresses this 'who is who in the Psalms?' problem. Within Old Testament scholarship the discipline of participant analysis has researched this question. The study breaks the research problem down in three subproblems, and aims to address them. Firstly, the reference data in Biblical Hebrew (BH) texts, specifically the Psalms, causes exegetical and theological problems for the reader. From a wealth of reference categories and data the right entities or participants needs to be identified. Secondly, a misbalance was observed on the one hand between the literary-rhetorical analysis and linguistic analysis of biblical poetry and between the study of prose and BH poetry on the other hand. Thirdly, the observed misbalance has affected the linguistic study of participant analysis in BH poetry, and the Psalms specifically. Besides studies conducted by scholars from the Eep Talstra Centre for Bible and Computer (ETCBC), De Regt and Runge, little to no attention has been paid to the linguistic study of participants in BH poetry.

It was observed that participant analysis as a form of discourse analysis studies coherence (the meaning relations between textual units), cohesion (the use of linguistic devices to link together textual units) and ambiguity (to whom or what is referred?). Participant analysis requires complex inference mechanisms. For resolving the correct coreference information to identify the correct entity, knowledge of the world is needed that cannot (always) be inferred from the (immediate) local or overall discourse. The study therefore acknowledged beforehand that participant analysis is pre-eminently subjective. The study's objective was to support participant analyses with more objective and explicit reference data.

A case study of how the exegetes in nine exegetical commentaries analyse Ps 75 demonstrates that they first establish semantic labels for the strophes and a genre for the entire text, and identify a variety of often contradicting entities with little to no basis in the psalm before the reference data is properly analysed. The commentaries have no systematic method for the analysis of reference data. It is therefore ambiguous how commentators retrieve their reference data, how they analyse it, identify entities and refer to them in their analyses. The methods and reference data are hardly reproducible and shareable; the interpretations and identifications of participants are thus more based on the author's views than that they have their basis in the text.

In relation to the theoretical framework of the dissertation, it was stated that participant analysis has a quantitative and qualitative component. The quantitative approach of participant analysis is the structured and reproducible gathering, labeling, identification, storage and easy accessibility and sharing of reference data. The qualitative approach is the interpretation and evaluation of the reference data and the tools that produced it. The methodological framework of Algorithmic Criticism is able to reconcile the quantitative approach of machine reading, understood as a coreference resolution task, with the qualitative approach of literary criticism and Old Testament exegesis.

Therefore, as a solution to the aforementioned research problems and to the absence of systematic participant analysis methods for Biblical Hebrew narrative and poetic texts, two computational methods were developed. The first machine reading method is a coreference annotation method and the second method is a deterministic coreference resolution algorithm called 'MiMi'. MiMi is the concatenation of *Mi Mi* in Biblical Hebrew and means 'Who? Who?' Both methods can meet the demands of a sound participant analysis method that starts with the linguistic analysis of language as system. A systematic, reproducible method that gives precedence to the syntactic analysis of features in BH poetic texts. Both methods are developed to support the interpretation process of participant analysis in Biblical Hebrew and not to replace the human interpreter.

Casting participant analysis for Biblical Hebrew as the interpretation of multiple coreference resolution operations with at its core Algorithmic Criticism to unify both the coreference annotation method, MiMi and their data products, with evaluation and exegetical interpretation, partitions the method into a sequence of systematic, well-defined, explainable and reproducible subtasks. Each method, the annotation method and MiMi, resolves coreference per Hebrew Bible chapter.

The annotation model first explicitly defines which linguistic reference data, i.e. 'mentions', and relations, i.e. 'coreference relations that are reflexive, symmetric and transitive', are considered for coreference resolution. The BH (chapter) text is obtained with Text-Fabric and transformed with the second step, mention detection, which is done either manually with annotation tools, or automatically with a context-free grammar. As a third step, the mentions are then analysed and (partially) resolved for coreference, either with manual linking with an annotation tool or automatically, sieve-based. Resolving mentions into classes means that they form an entity. This forming of entities is important for participant analysis. As a fourth step the coreference data and performance statistics are evaluated for their validity with either IAA analysis, or error analysis and mention detection analysis. A fifth and last step then, is to interpret what the meaning of the generated coreference data is for the text that is under study.

With the coreference annotation method a coreference-annotated corpus was created that also represents the Hebrew Bible genres. The corpus consists of the entire Psalms corpus (poetic), Genesis 1 and Numbers 8–10 (both narrative), and Isaiah 42

(prophetic). Ten Psalms chapters and three Numbers chapters were quantitatively and qualitatively analysed for inter-annotator agreement (IAA). The IAA statistics and analyses showed promising results. The insights from the results that the coreference annotation method brought forth were then integrated in MiMi. Though MiMi was initially optimised for the Psalms, it operates on all genres of the Hebrew Bible. With MiMi's coreference resolution rate of 57.1% for the Psalms and an even higher, average rate of 63.5% for the Hebrew Bible in general, an important beginning has been made for automated participant analysis.

Both coreference resolution methods demonstrate that it is possible to systematically, even automatically analyse Biblical Hebrew reference data for entities. It can be concluded that Biblical Hebrew makes use of a single reference system to refer to entities. There is no difference in methodology between the resolving of coreference data in poetic, narrative, or prophetic text.

From the evaluation of the data it became clear that the participant analysis task as it is understood in this dissertation is indeed a hard Natural Language Processing (NLP) and understanding task. The IAA measures for the Psalms and Numbers, MiMi's performance analyses and the IAA measures for the annotation of Ps 75, and MiMi's results made this clear. This is why working as much as possible in the Open Science way by applying the FAIR data principles – Findable, Accessible, Interoperable, and Reusable – is imperative for participant analysis as method. The NLP hardness of the participant analysis task demands that it should be done in comparison with other annotators or methods. The FAIRness of the methods, enable self-criticism and criticism by other researchers.

The computational methods developed, the coreference resolution annotation method and MiMi, enables Algorithmic Criticism of Biblical Hebrew text. With both coreference resolution methods the case study of Psalm 75 was revisited.

By using MiMi's data and the coreference annotations a 'minimum reading' and 'maximum reading' respectively of the text (of Ps 75) is established. Where MiMi 'reads' the text based on primarily syntactic features, the annotations contain coreference data based on syntactic features and additional semantic knowledge of the author. For the 'calibration' between the minimum reading, the maximum reading and the commentaries' interpretations, MiMi functions as critical algorithmic dialogue partner, as counterpart, for the coreference annotations. MiMi forces the interpreter into a self-critical dialogue in which deviations from the minimum reading – considered as added semantic information – demand an explanation. By calibrating between MiMi's data, the annotations and the interpretations of the commentaries, a well-informed interpretation of Ps 75 was enabled.

By working inside-out – from entities in Ps 75 as local discourse to entities in the corpus – it was more clear which steps were taken to resolve and interpret entities from within and outside the text. It was thereby also clear what the origin of the reference data was, which reference data was used for which interpretation. In other words: where MiMi's minimum reading of the text had flown over into interpretation

that went beyond the horizon of the text itself. By abstracting the textual data to mentions with identifiers and partitioning them in numbered classes, i.e. entities, and singletons, interpretation and entity identification issues both in the data and in the commentaries were laid bare. By working inside-out, the structured data allowed for searches in the corpus for similar entities and similar coreference chain introduction patterns for gaining a better understanding of the entities of Ps 75, and thus of its discourse.

The Algorithmic Criticism of Ps 75 thereby revealed that the identifications of king, prophet or priest have no basis in the text or corpus and are unnecessary for enhancing the understanding of the psalm. Algorithmic Criticism was able to separate God and human speech. In addition, Algorithmic Criticism revealed that, in contrast to what the commentaries claim, some aspects of Ps 75 remain unsolved. Unsolved issues are for example the unidentified speakers in vv. 7–9 and 10. The interpretation of the meaning of the data led to a categorisation of Ps 75 as an eschatological and universal judgement psalm.

The methods and the coreference data make the interpreter stay closer to the text which means that the interpretations of entities are less creative but do have a firm base in the text itself. By grounding the interpretations in reference data that is FAIR, produced by FAIR methods, acts of criticism and fruitful exegesis is stimulated. With the described methods 'the who is who question' can be answered in a more systematic way.